W 1108

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-306407

(43)Date of publication of application : 02.11.2001

(51)Int.Cl. G06F 12/16
G06F 3/06
G06F 13/00
G11B 20/10

(21)Application number : 2000-118012

(22)Date of filing : 19.04.2000

(71)Applicant : HITACHI LTD

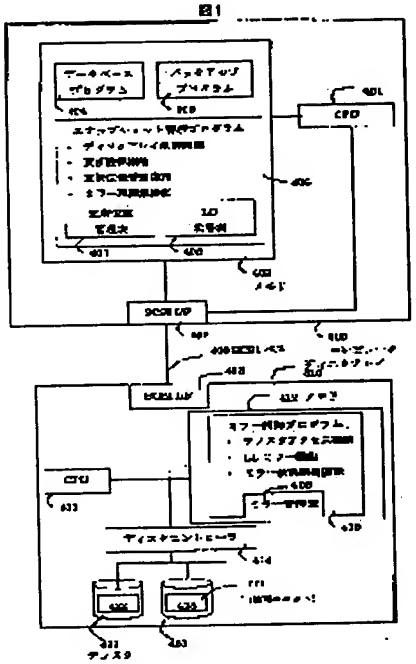
(72)Inventor : AJIMATSU YASUYUKI
MATSUNAMI NAOTO
MURAOKA KENJI
OEDA TAKASHI
YAGISAWA IKUYA

(54) METHOD FOR MANAGING SNAPSHOT AND COMPUTER SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To manage a snapshot without increasing the CPU load of a computer and the communication quantity of data due to the duplex processing of data writing by evading the concentration of load onto a specific medium in backup operation.

SOLUTION: When a snapshot is not taken, data writing from the computer to an external storage device 410 is duplicated and stored in respective areas of plural different storage media 420, 421. The duplication of data writing is performed by the external storage device 410. In the ease of taking a snapshot of external storage, one area out of the duplicated areas is provided as a storage area for a normal access and the other area is provided as a snapshot area. When the normal access storage area is updated during the storage of the snapshot, the computer allows the contents of the two areas to coincide with each other at the time of deleting the snapshot.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

W1108

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報**（Ａ）　(11)特許出願公開番号
特開2001－306407
（P2001－306407A）
(43)公開日　平成13年11月2日(2001.11.2)

(51)Int.Cl.[7]	識別記号	FI		テーマコード（参考）
Ｇ０６Ｆ　12/16	３１０	Ｇ０６Ｆ　12/16	３１０Ｍ	５Ｂ０１８
			３１０Ｊ	５Ｂ０６５
3/06	３０４	3/06	３０４Ｆ	５Ｂ０８３
	５４０		５４０	５Ｄ０４４
13/00	３０１	13/00	３０１Ｐ	

審査請求　未請求　請求項の数７　ＯＬ　（全 17 頁）　最終頁に続く

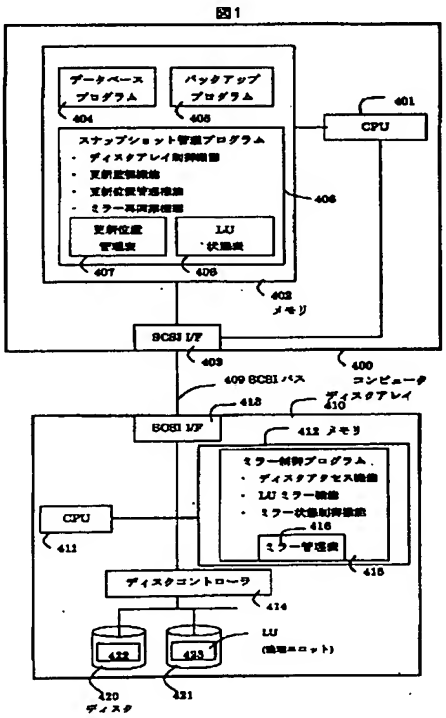
(21)出願番号　特願2000－118012(P2000－118012)
(22)出願日　平成12年４月19日(2000.4.19)

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　味松　康行
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(72)発明者　松並　直人
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(74)代理人　100078134
弁理士　武　顕次郎

最終頁に続く

(54)【発明の名称】　スナップショット管理方法及び計算機システム

(57)【要約】
【課題】　バックアップ中の特定媒体への負荷の集中を避け、データ書き込みの二重化処理に伴うコンピュータのＣＰＵ負荷やデータ通信量の増加を生じさせることなくスナップショットの管理を行う。
【解決手段】　スナップショットを取っていないときには、コンピュータから外部記憶装置へのデータ書き込みを、異なる複数の記憶媒体４２０、４２１上の領域に二重化して記憶する。データ書き込みの二重化は、外部記憶装置４１０で行う。外部記憶のスナップショットを取る際、二重化された領域のうち１つの領域を通常のアクセス用の記憶領域として提供し、もう１つの領域をスナップショットとして提供する。スナップショット保存中に通常のアクセス用の記憶領域が更新された場合、スナップショット削除時にコンピュータが、２つの領域の内容を一致させる。



【特許請求の範囲】

【請求項１】　複数の記憶媒体を持つ外部記憶装置を備える計算機システムにおけるスナップショット管理方法おいて、前記外部記憶装置は、内部の異なる２つの記憶媒体上の記憶領域をアクセス可能な独立な記憶領域としてコンピュータに提供すると共に、前記２つの記憶領域をグループとして定義し、グループ内の２つの記憶領域をアクセス可能な１つの仮想的な記憶領域としてコンピュータに提供し、前記仮想的な記憶領域に対してコンピュータがデータ書き込みを要求したとき、そのデータをグループ内の２つの記憶領域に書き込んで二重化し、コンピュータ内で動作するプログラムがスナップショットの取得を要求したとき、前記２つの記憶領域を独立した記憶領域としてアクセス可能とし、前記コンピュータ内で動作するプログラムは、スナップショット取得中におけるスナップショットを取得された記憶領域に対するデータの書き込みを検出し、データの書き込み先の位置を記録することを特徴とするスナップショット管理方法。

【請求項２】　前記外部記憶装置は、各記憶領域についてその状態を記憶し、コンピュータからの指示により指定された記憶領域の状態を変更し、記憶領域の状態に応じて、記憶領域へのアクセスの制限、グループ内の２つの記憶領域へのデータ書き込みの二重化の制御を行い、コンピュータ内で動作するプログラムは、記憶領域の状態を記憶し、外部記憶装置に記憶領域の状態を変更する指示を送信し、前記外部記憶装置は、スナップショット削除時に、記録された更新位置のデータを記憶領域間でコピーすることによりグループ内の２つの記憶領域の内容を一致させることを特徴とする請求項１記載のスナップショット管理方法。

【請求項３】　スナップショット削除時、コンピュータ内で動作するプログラムは、コピーする必要があるデータの位置を外部記憶装置に送信し、外部記憶装置は、その位置情報に基づいて領域間でデータコピーを行うことを特徴とする請求項２記載のスナップショット管理方法。

【請求項４】　コンピュータ内で動作するプログラムは、記憶領域の状態を一括管理し、スナップショット削除時に、記録された更新位置のデータを記憶領域間でコピーすることによりグループ内の２つの記憶領域の内容を一致させることを特徴とする請求項１記載のスナップショット管理方法。

【請求項５】　ネットワークで接続された別のコンピュータ上のプログラムが、スナップショットの取得及び削除を指示し、外部記憶装置が、オリジナルのデータにアクセスするコンピュータとは異なるコンピュータにスナップショットを提供することを特徴とする請求項４記載のスナップショット管理方法。

【請求項６】　コンピュータ内で動作するプログラムは、スナップショット取得後に、スナップショットを取得された記憶領域の内容を更新する際、更新前のデータを別の記憶領域に保存すると共に、保存先の位置を記録し、スナップショットを、コンピュータ上で動作する他のプログラムがアクセスできる仮想的な記憶領域として提供し、他のプログラムがスナップショットの読み出しを要求したとき、アクセス先がスナップショット取得後に更新されていれば、別の記憶領域に保存された更新前のデータを読み出し、アクセス先が更新されていなければ、外部記憶装置が提供する仮想的な記憶領域を構成する２つの記憶領域のうち、特定の１つからデータを読み出すことを特徴とする請求項１記載のスナップショット管理方法。

【請求項７】　複数の記憶媒体を持つ外部記憶装置を備える計算機システムにおいて、前記外部記憶装置は、内部の異なる２つの記憶媒体上の記憶領域をアクセス可能な独立な記憶領域としてコンピュータに提供すると共に、前記２つの記憶領域をグループとして定義し、グループ内の２つの記憶領域をアクセス可能な１つの仮想的な記憶領域としてコンピュータに提供する手段と、前記仮想的な記憶領域に対してコンピュータがデータ書き込みを要求したとき、そのデータをグループ内の２つの記憶領域に書き込んで二重化する手段と、コンピュータ内で動作するプログラムがスナップショットの取得を要求したとき、前記２つの記憶領域を独立した記憶領域としてアクセス可能とする手段とを備え、前記コンピュータは、その内部で動作するプログラムが、スナップショット取得中におけるスナップショットを取得された記憶領域に対するデータの書き込みを検出する手段と、データの書き込み先の位置を記録する手段とを備えることを特徴とする計算機システム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、スナップショット管理方法及び計算機システムに係り、特に、高可用性を求められる計算機システムにおける外部記憶をバックアップする際に必要な外部記憶のスナップショット管理方法、及び、この管理方法により管理される記憶装置を有する計算機システムに関する。

【０００２】

【従来の技術】一般に、ハードディスク等のコンピュータの外部記憶装置に記録されたデータは、装置の障害、ソフトウェアの欠陥、誤操作などによりデータを喪失した場合に、喪失したデータを回復できるように定期的に磁気テープ等にコピーして保存しておくこと(バックアップ)が必要である。その際、コピー作業中にデータが更新されるとコピーしたデータに不整合が生じるため、コピー作業中にデータが更新されないことを保証する必要がある。

【０００３】バックアップされるデータの更新を避けるためには、データにアクセスするバックアッププログラ

ム以外のプログラムを停止させればよいが、高可用性が要求されるシステムは、プログラムを長時間停止させることができない。このため、バックアップ中にプログラムがデータを更新することを妨げず、なおかつバックアップ開始時点でのデータの記憶イメージを保存する仕組みが必要である。

【０００４】通常、ある時点でのデータの記憶イメージをスナップショットと呼び、指定された時点のスナップショットを保存しつつデータの更新が可能な状態を提供する仕組みをスナップショット管理方法と呼ぶ。また、スナップショット管理方法によりスナップショットを保存することをスナップショットの取得と呼び、スナップショット取得の対象となったデータをオリジナルデータと呼ぶ。

【０００５】従来のスナップショット管理方法は、更新前データの保存、コンピュータによるデータの二重化、あるいは、外部記憶装置によるデータ二重化によって実現されている。以下、これらの方法による従来のスナップショット管理方法について説明する。

【０００６】（１）更新前データの保存による方法

この方法に関する従来技術として、例えば、米国特許第５６４９１５２号明細書に記載された技術が知られている。この従来技術は、あるデータのスナップショットを取得するとき、スナップショット取得時点以後にオリジナルデータを更新する場合、更新前の記憶内容を別の記憶領域に保存するというものである。スナップショットは、論理的にはオリジナルデータとは独立な別データとしてアクセスされるが、スナップショット取得時点以後にオリジナルデータが更新されていない部分について、スナップショットはオリジナルデータと記憶領域を共有する。オリジナルデータが更新された部分は、別領域に保存された更新前の記憶内容が参照される。また、この場合、スナップショットは読み出し専用である。

【０００７】（２）コンピュータによるデータの二重化による方法

コンピュータによるデータの二重化による方法は、スナップショットを取得していない通常の状態において、コンピュータ上のプログラムが全てのデータを２つの記憶領域に二重化（ミラー）するというものである。スナップショットを取得するとき、コンピュータ上のプログラムは、二重化の処理を停止して２つの記憶領域を独立な領域に分離し、１つの領域をオリジナルデータ、もう１つの領域をスナップショットとして提供する。コンピュータ上のプログラムがデータを二重化する方法は、例えば、米国特許５０５１８８７号明細書等に記載されて知られている。

【０００８】（３）外部記憶装置によるデータの二重化による方法

外部記憶装置によるデータの二重化による方法は、例えば、米国特許５８４５２９５号明細書に示されているように、前述した（２）の方法において、コンピュータ内のプログラムが行ったスナップショット管理を外部記憶装置内で行うというものである。スナップショットを管理する機能は、全て外部記憶装置内のスナップショット管理プログラムにある。

【０００９】

【発明が解決しようとする課題】前述した更新前データの保存による従来技術の方法は、スナップショット取得後、オリジナルデータが更新されていない部分について、スナップショットに対するアクセスも、オリジナルデータに対するアクセスも同一の記憶領域を参照する。このため、この従来技術は、バックアップ中、特定の記憶媒体にアクセスが集中し、ディスク入出力の性能が低下するという問題点を有している。この結果、例えば、計算機システムの特定のディスク装置にアクセスが集中し、データベースプログラムやバックアッププログラムがそのディスク装置のＬＵ内のデータを読み出す速度が低下してしまうことになる。

【００１０】また、この従来技術は、オリジナルデータへのアクセスとスナップショットへのアクセスとのどちらも必ずスナップショット管理プログラムを経由しなければならないため、バックアップ中、スナップショット管理プログラムを実行するコンピュータの負荷が増大し、同一コンピュータ内で動作するデータベースプログラム等の実行速度が低下するという問題点を有している。この結果、特に、大量のデータをバックアップする際等に、バックアップ終了までの長時間にわたり性能を低下させてしまうことになる。

【００１１】また、前述したコンピュータによるデータの二重化による従来技術の方法は、スナップショットを取得していない状態でのデータ書き込みにおいて、１回のデータ書き込みに対してスナップショット管理プログラムが２つのＬＵにデータを書き込む必要があるため、スナップショット管理をしないシステムと比較して、２倍の書き込み処理を実行する必要がある。従って、この従来技術は、データの書き込みを行うコンピュータのＣＰＵ負荷や、コンピュータと外部記憶装置とを接続する通信路のデータ通信量、外部記憶装置のディスクコントローラ負荷が増大し、スナップショット管理をしないシステムと比較してアプリケーションプログラムの実行速度を低下させてしまうという問題点を生じる。特に、データベースのレプリケーション等、大量のデータ更新が伴う処理において性能の低下が顕著となる。

【００１２】さらに、外部記憶装置によるデータの二重化による従来技術の方法は、スナップショット管理に必要な全ての処理を外部記憶装置内に実装するため、外部記憶装置の制御プログラムが複雑になり、複雑化に対応して開発期間が長くなり、価格も上昇するため、このような外部記憶装置を用いる構成のシステムは高価になってしまうという問題点を有している。

【００１３】前述したように、従来技術は、バックアップ時の特定記憶媒体へのアクセス集中及びＣＰＵ負荷の増大、通常運用時のＣＰＵ、通信路及び外部記憶装置のコントローラの負荷の増大、あるいは、外部記憶装置の複雑化と高価格化という問題点を有している。

【００１４】本発明の目的は、前述した従来技術の問題点を解決し、外部記憶のスナップショットを取得した状態において、外部記憶装置の特定の記憶媒体への負荷集中やコンピュータのＣＰＵ負荷の増大を防ぐことができ、外部記憶のスナップショットを取得していない状態において、コンピュータのＣＰＵ負荷やデータ通信路の通信量及び外部記憶装置の負荷の低いスナップショット管理方法を提供することにある。

【００１５】また、本発明の目的は、比較的単純な機能だけを持つ安価な外部記憶装置を用いて、外部記憶のスナップショット管理を行う計算機システムを提供することにある。

【００１６】

【課題を解決するための手段】本発明によれば、複数の記憶媒体を持つ外部記憶装置を備える計算機システムにおけるスナップショット管理方法おいて、前記外部記憶装置が、内部の異なる２つの記憶媒体上の記憶領域をアクセス可能な独立な記憶領域としてコンピュータに提供すると共に、前記２つの記憶領域をグループとして定義し、グループ内の２つの記憶領域をアクセス可能な１つの仮想的な記憶領域としてコンピュータに提供し、前記仮想的な記憶領域に対してコンピュータがデータ書き込みを要求したとき、そのデータをグループ内の２つの記憶領域に書き込んで二重化し、コンピュータ内で動作するプログラムがスナップショットの取得を要求したとき、前記２つの記憶領域を独立した記憶領域としてアクセス可能とし、前記コンピュータ内で動作するプログラムが、スナップショット取得中におけるスナップショットを取得された記憶領域に対するデータの書き込みを検出し、データの書き込み先の位置を記録することにより達成される。

【００１７】

【発明の実施の形態】以下、本発明によるスナップショット管理方法及び計算機システムの実施形態を図面により詳細に説明する。

【００１８】図１は本発明の第１の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図、図２はミラー管理表の構成を示す図、図３は更新位置管理表の構成を示す図、図４はMode Select及びModeSenseで使用するパラメータページの内容を説明する図、図５はスナップショット管理プログラムのデータ書き込み時の動作を説明するフローチャート、図６はコンピュータからＷＲＩＴＥコマンドを受信したミラー制御プログラムの動作を説明するフローチャート、図７はスナップショット管理プログラムのミラー再同期機能の動作を説明するフローチャートである。図１～図３において、４００はコンピュータ、４０１、４１１はＣＰＵ、４０１、４１２はメモリ、４０３、４１３はＳＣＳＩインタフェース、４０４はデータベースプログラム、４０５はバックアッププログラム、４０６はスナップショットプログラム、４０７は更新位置管理表、４０８はＬＵ状態表、４０９はＳＣＳＩバス、４１０はディスクアレイ、４１４はディスクコントローラ、４１５はミラー制御プログラム、４１６はミラー管理表、４２０、４２１はディスク装置、４２２、４２３はＬＵである。本発明の第１の実施形態は、データベースプログラムがアクセスしている記憶領域を、データベース処理と並行してオンラインバックアップする場合に、本発明によりスナップショットを管理する例である。

【００１９】図１に示す計算機システムは、コンピュータ４００とディスクアレイ４１０とがＳＣＳＩインタフェース４０３、４１３を介してＳＣＳＩバス４０９により接続されて構成されている。コンピュータ４００内のメモリ４０２には、データベースプログラム４０４、バックアッププログラム４０５、スナップショット管理プログラム４０６が格納され、コンピュータを制御するＣＰＵ４０１によって実行される。ディスクアレイ４１０には、ディスクコントローラ４１４によって制御されるディスク装置４２０、４２１が設けられ、また、メモリ４１２内には、ミラー制御プログラム４１２が格納され、ＣＰＵ４１１によって実行される。各ディスク内の記憶領域は、ＳＣＳＩの論理ユニット（ＬＵ）４２２、４２３としてコンピュータ４００からアクセスされる。

【００２０】データベースプログラム４０４は、実行中にＬＵ４２２にアクセスし、また、ユーザからの指示によりディスク上のデータの更新を停止、再開する機能を持つ。バックアッププログラム４０５は、ユーザからの指示により、ＬＵ４２３からバックアップのためのデータ読み出しを行う。スナップショット管理プログラム４０６は、デバイスドライバやボリュームマネージャのようなミドルウェアであり、データベースプログラム４０４がディスクアレイに出すディスクアクセス要求の全てを自スナップショット管理プログラム４０６を経由してディスクアレイ４１０に発行する。そして、スナップショット管理プログラム４０６は、ディスクアレイ制御機能、更新監視機能、更新位置管理機能、ミラー再同期機能を有すると共に、更新位置管理表４０７、ＬＵ状態表４０８を有する。

【００２１】ディスクアレイ４１０のミラー制御プログラム４１５は、コンピュータ４００からの要求に応じてディスクコントローラ４１４にディスクアクセスを指示するディスクアクセス機能の外に、１つのＬＵに対する更新を二重化して予め指定された別のＬＵにも適用し、２つのＬＵに同一の内容を書き込むＬＵミラー機能を持つ。図示例の本発明の本実施形態は、ＬＵ４２２をＬＵ

４２３に二重化する機能を有する。これらは従来のディスクアレイが持つ機能である。また、ミラー制御プログラム４１５は、ＬＵミラー機能を無効化することにより、二重化された２つのＬＵをそれぞれアクセス可能な独立のＬＵとするミラー状態制御機能を有する。

【００２２】二重化される各ＬＵのミラー先の定義と現在の状態とは、ミラー管理表４１６に記録される。ミラー管理表４１６は、図２にその構成を示すように、二重化される各ＬＵについて、そのミラー先ＬＵと状態とが記録される。状態は、２つのＬＵの内容が一致している「同期」、それぞれが独立なＬＵとしてアクセスされる「スナップショット中」、「スナップショット中」から「同期」に移るためにＬＵ間で一致しないデータをコピーしている「再同期中」の３種類がある。ＬＵミラー機能は状態が「同期」、「再同期中」のＬＵに対して適用される。ミラー管理表４１６において、ミラー先ＬＵを複数にすることにより、３重以上の多重化構成とすることもできる。

【００２３】スナップショット管理プログラム４０６は、後述するように、ＳＣＳＩのModeSense,Mode Selectコマンドを発行するディスクアレイ制御機能、スナップショット取得後にデータベースプログラムが要求するオリジナルデータの更新を検出する更新監視機能、その位置を記録する更新位置管理機能、及び、スナップショット削除時にオリジナルデータの更新部分をミラー先ＬＵにコピーするミラー再同期機能を有している。更新監視機能及び更新位置管理機能は、状態が「スナップショット中」または「再同期中」のＬＵに対して、また、ミラー再同期機能は、状態が「再同期中」のＬＵに対してのみ適用される。

【００２４】各ＬＵの現在の状態を記録するＬＵ状態表４０８は、図２に示すディスクアレイのミラー管理表４１６からミラー先ＬＵの定義を省略した構成を有する。ＬＵ状態表４０８は、スナップショット管理プログラム４０６の起動時に、アクセス可能な全てのＬＵについてMode Senseコマンドを発行し、ディスクアレイ４１０から各ＬＵの状態を得ることにより作成される。また、更新位置管理表４０７には、スナップショット取得後のオリジナルデータの更新位置が記録される。更新位置管理表には、図３に示すように、更新されたオリジナルデータを記憶するＬＵ、更新された位置の最初のＬＢＡ、最後のＬＢＡが記録される。図３に示す例の場合、ＬＵ２のＬＢＡ０～１５及びＬＢＡ２５６～５１１、ＬＵ３のＬＢＡ８～３１の部分がスナップショット取得後に更新されていることが示されている。更新位置管理表４０７の記録は、二重化のペアとして定義された２つのＬＵ間で内容が一致しない位置を表す。

【００２５】次に、前述したように構成される本発明第１の実施形態による計算機システムにおけるコンピュータからのディスクアレイの制御について説明する。

【００２６】ＬＵミラー機能の有効／無効化やミラー先ＬＵに対するアクセスの可否等のディスクアレイのミラー制御プログラムの各ＬＵに対する動作は、全てミラー管理表に記録されているＬＵの状態によって決定される。ミラー制御プログラム４１５は、ミラー管理表４１６の状態記録をスナップショット管理プログラム４０６が発行するＳＣＳＩ Mode Selectコマンドで指定された状態に変更する。また、ミラー制御プログラム４１５は、ミラー管理表４１６の内容を要求するＳＣＳＩMode Sense コマンドを受信すると、指定されたＬＵの現在の状態記録を返送する。Mode Select及びModeSenseで使用するパラメータページの内容を図４に示している。パラメータページは、２バイトからなり、先頭にはページコード(00h)が指定されており、２バイト目は指定されたＬＵの状態を表している。Mode Select は、「同期」、「スナップショット中」、「再同期中」に応じて、それぞれ、０、１、２を指定する。また、Mode Senseは、現在の状態を返送するが、ModeSenseコマンドで指定したＬＵが二重化されていなかったり、ミラー先ＬＵとして定義されている場合、「未定義」として４が指定される。

【００２７】次に、スナップショットの取得について説明する。ＬＵ４２２のスナップショットを取得するには、以下に説明するような処理を行う。

【００２８】まず、ユーザが一時的にデータベースの更新を停止し、バックアップするデータの整合性を保証する。次に、スナップショット管理プログラムにＬＵ４２２のスナップショットの取得を指示する。スナップショット管理プログラム４０６は、ＬＵ状態表４０８を参照し、ＬＵ４２２の状態が「同期」であることを確認するが、ＬＵ状態表にＬＵ４２２の記録がない場合や、状態が「同期」でない場合、処理を中止する。

【００２９】ＬＵ４２２の状態が「同期」である場合、スナップショット管理プログラム４０６はＬＵ状態表４０８に記録された状態を「同期」から「スナップショット中」に変更して更新監視機能及び更新位置管理機能を有効化すると共に、ディスクアレイ４１０内のミラー管理表４１６に記録されたＬＵ４２２の状態を「スナップショット中」に変更するためにMode Selectコマンドを発行する。ディスクアレイ４１０のミラー制御プログラム４１５は、Mode Selectコマンドを受信すると、ミラー管理表４１６に記録されたＬＵ４２２の状態を「スナップショット中」に変更する。その後、ユーザは、データベースの更新を再開する。

【００３０】データの読み出し、書き込みの制御として後述するように、ミラー制御プログラム４１５の各ＬＵに対する動作はＬＵの状態により決定される。従って、この状態変更により、ミラー先ＬＵであるＬＵ４２３が独立したＬＵとしてコンピュータからアクセス可能となり、また、ＬＵ４２２に対する更新がＬＵ４２３に反映

されなくなるため、バックアッププログラム４０５は、ＬＵ４２３をＬＵ４２２のスナップショットとして利用することが可能となる。

【００３１】次に、スナップショットの削除について説明する。ＬＵ４２２のスナップショットを削除するには、以下に説明するような処理を行う。

【００３２】まず、ユーザがスナップショット管理プログラム４０６にＬＵ４２２のスナップショットの削除を指示する。スナップショット管理プログラム４０６は、ＬＵ状態表４０８を参照し、ＬＵ４２２の状態が「スナップショット中」であることを確認するが、ＬＵ状態表４０８にＬＵ４２２の記録がない場合や、状態が「スナップショット中」でない場合処理を中止する。

【００３３】ＬＵ４２２の状態が「スナップショット中」である場合、スナップショット管理プログラム４０６は、ディスクアレイ４１０内のミラー管理表４１６に記録されたＬＵ４２２の状態を「再同期中」に変更するためにMode Select コマンドを発行する。ディスクアレイのミラー制御プログラム４１５は、Mode Select コマンドを受信すると、ＬＵ４２２の状態を「再同期中」に変更する。この状態変更により、ＬＵ４２３をアクセス先とするコンピュータからのアクセスが不可能となり、また、ＬＵ４２２の更新をＬＵ４２３に反映するＬＵミラー機能が有効化される。

【００３４】次に、スナップショット管理プログラム４０６は、ＬＵ状態表４０８に記録されたＬＵ４２２の状態を「再同期中」に変更し、ミラー再同期機能を有効化する。ミラー再同期機能によるデータコピー処理の詳細については後述する。スナップショット管理プログラム４０６は、データコピー処理が完了すると、再びModeSelect コマンドを発行してディスクアレイ４１０内のミラー管理表４１６に記録されたＬＵ４２２の状態を「同期」に変更し、ＬＵ状態表４０８に記録された状態を「同期」に変更する。以上により、スナップショット管理プログラムのＬＵ４２２に対する更新監視機能、更新位置管理機能、ミラー再同期機能が無効化され、スナップショットの削除が完了する。

【００３５】次に、データの読み出しの処理、すなわち、コンピュータ上のデータベースプログラム４０４、バックアッププログラム４０５が、それぞれディスクアレイのＬＵ４２２、４２３内のデータを読み出すときの処理手順について説明する。

【００３６】コンピュータ４００内のプログラムからの入出力要求は、スナップショット管理プログラム４０６を経由するが、読み出しの場合、スナップショット管理プログラム４０６は何もせずにＳＣＳＩのＲＥＡＤコマンドをディスクアレイに転送する。ディスクアレイのミラー制御プログラム４１５は、ＲＥＡＤコマンドを受信すると、指定されたＬＵからデータを読み出し、データとステータスとをコンピュータ４００に返送する。但し、ミラー制御プログラム４１５は、指定されたＬＵの状態が「同期」であれば、指定されたＬＵとそのミラー先ＬＵの内容とが一致しているので、両者のいずれかから読み出すことにより負荷を分散させる。また、ミラー制御プログラム４１５は、指定されたＬＵがミラー先ＬＵで、かつ、状態が「スナップショット中」でない場合、アクセスを拒否し、エラーステータスを返送する。なお、スナップショット中でないミラー先ＬＵへのアクセスは、書き込みの場合も含めてスナップショット管理プログラム４０６内で拒否するようにしてもよい。

【００３７】次に、データの書き込みの処理、すなわち、コンピュータ上のデータベースプログラム４０４、バックアッププログラム４０５が、それぞれディスクアレイのＬＵ４２２、４２３内の記憶内容を更新するときの処理動作を図５に示すフローを参照して説明する。

【００３８】（１）スナップショット管理プログラム４０６は、プログラム４０４、４０５からの更新要求を受け取ると、ＬＵ状態表４０８を参照し、アクセス先ＬＵの状態が「スナップショット中」であるか否かを確認する（ステップ８００、８０１）。

【００３９】（２）ステップ８０１での判定が「スナップショット中」であった場合、アクセス先の位置がすでに更新位置管理表４０７に記録されているか否かを調べ、記録されていなければ、更新位置管理表４０７に位置を記録して、ＳＣＳＩのＷＲＩＴＥコマンドを発行する。また、記録されていれば、そのままＳＣＳＩのＷＲＩＴＥコマンドを発行する（ステップ８０２、８０３、８０７）。

【００４０】（３）ステップ８０１での判定で、ＬＵの状態が「スナップショット中」でなかった場合、さらに「再同期中」か否かを確認し、もし「再同期中」でなければアクセス先ＬＵの状態は「同期」であるから、そのままＷＲＩＴＥコマンドを発行する（ステップ８０４、８０７）。

【００４１】（４）ステップ８０４での判定で、「再同期中」であれば、さらにアクセス位置が更新位置管理表４０７に記録されているか否かを調べ、記録がなければ、そのままＷＲＩＴＥコマンドを発行し、更新位置の記録があれば、表から該当位置の記録を削除してからＷＲＩＴＥコマンドを発行する（ステップ８０５～８０７）。

【００４２】前述において、更新位置管理表４０７の記録の追加・削除は、単純なエントリの追加・削除ではなく、追加の場合、同一位置を含む複数の記録が存在しないように、削除の場合、削除する位置以外の記録を削除しないように更新位置管理表４０７の修正を行う。例えば、更新位置管理表４０７が図３により説明したような内容を持ち、ＬＵ３のＬＢＡ＝１０～１１を削除する場合、表に記録されたＬＵ３の更新位置であるＬＢＡ＝８～３１を削除し、新たにＬＢＡ＝８～９及びＬＢＡ＝１

２～３１のエントリを追加する。

【００４３】次に、コンピュータからＷＲＩＴＥコマンドを受信したディスクアレイのミラー制御プログラム４１５の動作を、図６に示すフローを参照して説明する。

【００４４】（１）ＷＲＩＴＥコマンドを受信したミラー制御プログラム４１５は、まず、ミラー管理表を参照してアクセス先ＬＵがミラー先ＬＵとして記録されているか否かを調べる（ステップ９００、９０１）。

【００４５】（２）ステップ９０１の判定で、アクセス先がミラー先ＬＵであれば、状態が「スナップショット中」であるか否かを調べ、スナップショット中であれば指定されたＬＵにデータを書き込み、そうでなければ独立したＬＵとしてアクセスできないので、書き込み失敗のステータスを返送する（ステップ９０５、９０４、９０６）。

【００４６】（３）ステップ９０１の判定で、アクセス先ＬＵがミラー先ＬＵでない場合、状態が「スナップショット中」か否かを確認し、スナップショット中でなければミラー先ＬＵと指定されたＬＵにデータを書き込み、そうでなければ指定されたＬＵにのみデータを書き込んで、コンピュータにコマンド実行のステータスを返送する（ステップ９０２～９０６）。

【００４７】前述したスナップショット管理プログラム４０６の処理において、更新先のＬＵの状態が「再同期中」でアクセス位置が更新位置管理表４０７に記録されている場合に、更新位置管理表４０７の記録を削除する理由は、後述するスナップショット削除時のＬＵ再同期処理で実行されるデータコピー処理の無駄を無くすためである。「再同期中」の状態のＬＵに対する更新において、ミラー制御プログラム４１５のＬＵミラー機能（図６のステップ９０３）が有効であるため、データ更新により更新部分のＬＵの内容が一致し、再同期のデータコピー処理は必要がなくなる。再同期のデータコピーは、次に説明するように、更新位置管理表４０７に記録された位置について実行されるため、更新により内容が一致した位置の記録は更新位置管理表４０７から削除する。

【００４８】次に、スナップショット削除の際のデータコピーの処理について説明する。スナップショット削除の処理として前述で説明したように、スナップショットを削除する際に、二重化のペアとして定義された２つのＬＵで記憶内容が一致しない部分についてオリジナルデータをミラー先ＬＵにコピーする処理が必要となる。この処理を行うミラー再同期機能の動作を図７に示すフローを参照して説明する。スナップショット管理プログラムのミラー再同期機能は、他の機能と並行に動作する。

【００４９】（１）ミラー再同期機能は、処理を開始すると、まず、更新位置管理表４０７に記録があるか否かを調べ、もし、記録がなければ処理完了としてここでの処理を終了する（ステップ１０００～１００２）。

【００５０】（２）ステップ１００１で、更新位置管理表４０７に記録があった場合、記録された位置の更新済みデータを１つ読み出す。この読み出しは、ディスクアレイにＳＣＳＩのＲＥＡＤコマンドを発行することにより行われる（ステップ１００３）。

【００５１】（３）ミラー再同期機能は、ステップ１００３でのＲＥＡＤコマンドの発行に対してディスクアレイからデータが返送されてきたら、もう一度更新位置管理表４０７を参照し、読み出した位置の記録が削除されていないか否かを確認する。このような削除は、ＳＣＳＩＲＥＡＤコマンドを発行した直後にデータベースプログラムが同一位置の記録を更新した場合に発生する（ステップ１００４）。

【００５２】（４）ステップ１００４の判定で、読み出した全ての位置の記録が削除されている場合、更新により新しいデータが書き込まれて２つのＬＵの内容が一致しているのでデータコピーの必要がなくステップ１００１の処理に戻り、読み出したデータの一部または全部の位置情報が更新位置管理表４０７に残っている場合、記録されている位置のデータを読み出した位置に書き戻す。ＬＵの状態が「再同期中」の場合、データの書き込み処理として前述で説明したように、ミラー制御プログラムは、更新をミラー先ＬＵにも反映させる。これにより、２つのＬＵの内容を一致させることができる。（ステップ１００５）。

【００５３】（５）最後に、ミラー再同期機能は、更新位置管理表４０７の記録を削除し、ステップ１００１の処理に戻る（ステップ１００６）。

【００５４】前述した本発明の第１の実施形態は、コピーが必要なデータの位置をコンピュータ４００内の更新位置管理表４０７にのみ記録することとしているため、コンピュータ４００の障害により位置情報が失われる可能性がある。しかし、その場合、全てのオリジナルデータをミラー先にコピーすることにより、更新内容を失うことなく記憶内容を一致させることができる。

【００５５】前述した本発明の第１の実施形態によれば、データベースプログラムがアクセスするＬＵと、そのスナップショットとして使用されるミラー先ＬＵが異なるディスク上にあるため、バックアッププログラムとデータベースプログラムとがそれぞれ発生するディスクアクセスの負荷を分散させることができ、バックアップ中のデータベースプログラムのディスクアクセス速度の低下を軽減することができる。

【００５６】すなわち、いま、２つのプログラムが領域内を均等にアクセスすると仮定し、スナップショット取得後に更新されたデータベースの割合をｐ、データベースプログラムが発生するディスクアクセス負荷をＤ、バックアッププログラムが発生するディスクアクセス負荷をＢとする。すると、更新前データの保存による従来技術の方法の場合、更新されていない部分のデータは、バックアッププログラムもオリジナルデータと同じ記憶領

域を参照するため、ディスク装置４２０へのアクセス負荷はDから（D＋B（1－p））に増大するが、前述した本発明の第１の実施形態の場合、データベースアクセスとバックアップアクセスとは記憶領域を共有しないため負荷はDのままである。

【００５７】また、前述した本発明の第１の実施形態は、スナップショットを取っていないときにデータ書き込みを二重化するが、二重化処理が、ディスクアレイ内のミラー制御プログラムで実行されているため、コンピュータは１回のデータ書き込みにつき、ＳＣＳＩ　ＷＲＩＴＥコマンドを１回送信するだけでよい。このため、コンピュータによるデータの二重化による従来の方法と比べ、Ｉ／Ｏ処理のためのコンピュータのＣＰＵ負荷とＳＣＳＩバスの通信量とを低減することができ、データベースプログラムの実行速度を向上させることができる。

【００５８】さらに、前述した本発明の第１の実施形態で用いるディスクアレイは、スナップショット中の更新位置管理機能やスナップショット削除時のＬＵの再同期機能を持たないため、外部記憶装置によるデータ二重化による従来技術の方法に比較して、単純な機能のみを提供する安価なディスクアレイを用いてシステムを構成することが可能となる。

【００５９】前述した本発明の第１の実施形態の拡張形態として、ミラー制御プログラムに機能を追加し、ミラー削除時のコンピュータ内の処理を単純化することも考えられる。この場合、指定された位置のＬＵ間のデータコピーを行う機能をミラー制御プログラム４１５に追加することにより、スナップショット管理プログラムは実際にデータの読み出し、書き込みを行わずにコピーすべき位置をミラー制御プログラムに指示するだけで内容を一致させることができる。その場合、図７のステップ１００３～１００５の代わりに、コピー位置を指示するための予め定義されたＳＣＳＩコマンドを発行するステップを設ければよい。このコマンドとしては、ステップ１００３のＲＥＡＤコマンドと同じアクセス位置を指定し、受信したミラー制御プログラムは、ディスクアレイ内で指定された位置のデータをミラー先ＬＵにコピーするコマンドであればよい。これにより、データコピー処理に伴うコンピュータのＣＰＵ負荷やＳＣＳＩバスの通信量の増大を低減することができる。

【００６０】図８は本発明の第２の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図、図９はミラー定義表の構成を示す図、図１０はＬＵ状態表の構成を示す図、図１１はModeSense に対して返すパラメータページの内容を説明する図であり、以下、これらの図を参照して本発明の第２の実施形態について説明する。図８～図１１において、１１０８、１１１６はミラー定義表であり、他の符号は、図１の場合と同一である。この本発明の第２の実施形態は、前述で説明した第１の実施形態と同様にスナップショットを管理するものであるが、第１実施形態において、コンピュータ及び外部記憶装置の双方において行われていたＬＵの状態管理をコンピュータのみで行い、スナップショット管理に関して外部記憶装置をステートレス化したものである。

【００６１】図８に示す本発明の第２の実施形態において、ディスクアレイ４１０内のミラー制御プログラム４１５は、二重化された２つのＬＵを論理的な１つのＬＵとして提供するＬＵ仮想化機能を有する。ＬＵ仮想化機能で扱われるＬＵ間の関係は、図９に示すようなミラー定義表１１１６に記録される。このミラー定義表１１１６は、３つのＬＵを１組としたＬＵ間の関係を表し、オリジナルデータを持つミラー元ＬＵ、その内容を二重化したミラー先ＬＵ、これらを論理的に１つのＬＵとして扱うミラーＬＵを定義している。３つのＬＵは、全て外部からアクセス可能である。また、ミラー元ＬＵとミラー先ＬＵとの内容が一致するように保つ責任は、コンピュータ４００で動作するスナップショット管理プログラム４０６にある。

【００６２】スナップショット管理プログラム４０６は、ディスクアレイ４１０が提供するＬＵを仮想ＬＵとして他のプログラムに提供する。仮想ＬＵは、実際のＬＵと同様に他のプログラムからアクセスされる。プログラムが仮想ＬＵにアクセスしたときにアクセス先となる実際のＬＵは、仮想ＬＵの状態により、スナップショット管理プログラムが切り替える（ＬＵ仮想化機能）。仮想ＬＵのアクセス先及び状態は、図１０に示すようなＬＵ状態表４０８に記録される。実際にアクセス先となるＬＵは、読み出しの場合と書き込みの場合とで別々に定義される。ＬＵ対応表４０８に記録される状態は、第１実施形態におけるＬＵ状態表と同様であるが、それ以外に「未定義」の状態があり、これはその仮想ＬＵに対するアクセスができない状態を表す。

【００６３】更新位置管理表４０７は、図３により説明した第１の実施形態のものと同様な構造を持つが、変更されたＬＵは、実際のＬＵではなく仮想ＬＵが記録される。また、スナップショット管理プログラムの更新監視機能、更新位置管理機能、ミラー再同期機能がそれぞれ監視、管理、アクセスする対象ＬＵも仮想ＬＵである。各機能は、第１の実施形態の場合と同様に仮想ＬＵの状態に応じて有効化、無効化がなされる。「未定義」のＬＵに対して、全ての機能は無効化される。

【００６４】また、スナップショット管理プログラム４０６は、ディスクアレイ内のミラー定義表１１１６と同一内容のミラー定義表１１０８を持つ。このミラー定義表は、スナップショット管理プログラムの起動時に、全てのアクセス可能なＬＵに対してＳＣＳＩ Mode Senseコマンドを発行することにより、ディスクアレイ４１０からミラー定義表１１１６の情報を得て作成される。デ

ィスクアレイ４１０は、Mode Senseに対して図１１に示すような構造を持つ４バイトのパラメータページを返送し、各ＬＵについてミラー定義表１１１６の内容を知らせる。

【００６５】このパラメータページは、先頭の１バイト目にはページコード０が入り、ModeSenseで指定されたＬＵがミラーＬＵであれば２バイト目には“０”、そうでなければ“１”が入る。また、種別が“０”のとき、３バイト目、４バイト目にはそれぞれミラー元ＬＵ、ミラー先ＬＵが入る。ミラー定義表１１０８を作成する場合、種別が“０”のＬＵに対してエントリを作成し、ミラー元ＬＵ、ミラー先ＬＵを記録する。このミラー定義表１１０８の作成は、Mode Senseを用いた自動作成である必要はなく、ユーザが手動で定義した内容ファイルをスナップショット管理プログラム４０６が起動時に読み込んで作成してもよい。

【００６６】初期状態において、ＬＵ状態表４０８のデータベースプログラム４０４がアクセスする仮想ＬＵの状態は「同期」、ＲＥＡＤ先及びＷＲＩＴＥ先ＬＵはいずれもミラー定義表１１０８のミラーＬＵが記録され、バックアッププログラム４０５がアクセスする仮想ＬＵの状態は「未定義」が記録されている。

【００６７】次に、前述のように構成される本発明の第２の実施形態におけるコンピュータ上のプログラムのディスクアクセスについて説明する。

【００６８】コンピュータ上のプログラムからのディスクアクセスは、スナップショット管理プログラム４０６が提供する仮想ＬＵに対して要求される。スナップショット管理プログラム４０６は、要求を受け取ると、ＬＵ状態表４０８を参照し、指定された仮想ＬＵの状態が「未定義」でないか否かを確認し、「未定義」であればアクセスを拒否する。

【００６９】次に、スナップショット管理プログラム４０６は、指定された仮想ＬＵのＲＥＡＤ先ＬＵ、ＷＲＩＴＥ先ＬＵをＬＵ状態表４０８で調べ、発行するコマンドのアクセス先となる実際のＬＵを決定する。例えば、図１０に示す状態において、仮想ＬＵ０にアクセスすると、スナップショット管理プログラム４０６は、読み出し要求、書き込み要求共に、ＬＵ０をターゲットとするＳＣＳＩコマンドを発行する。また、仮想ＬＵ４に読み出し、書き込みを要求すると、スナップショット管理プログラム４０６は、それぞれＬＵ７、ＬＵ６にコマンドを発行する。スナップショット管理プログラム４０６は、ＬＵ状態表を書き換えることにより、各プログラムがアクセスする仮想ＬＵを固定したままで、アクセス先のＬＵを切り替えることができる。

【００７０】アクセスコマンドを受信したディスクアレイのミラー制御プログラム４１５は、ミラー定義表１１１６を参照して対応するＬＵへのアクセスを実行する。例えば、図９に示すように定義されているＬＵ０に対するアクセスを受信すると、ＬＵミラー制御プログラム４１５は、ミラー定義表を参照してＬＵ０が２つのＬＵからなるミラーＬＵであると判断し、更新ではＬＵミラー機能によりＬＵ１とＬＵ２に同じ内容を書き込み、読み出しではＬＵ１またはＬＵ２からデータを読み出す。ＬＵ１やＬＵ２にアクセスがあった場合、これらはミラーＬＵではないとして、それぞれＬＵ１またはＬＵ２にのみアクセスする。

【００７１】次に、本発明の第２の実施形態におけるスナップショットの取得、削除について説明する。前述した本発明の第１の実施形態は、スナップショットの取得・削除の際に、ディスクアレイに対してMode Select コマンドを発行してスナップショットを管理したが、第２の実施形態は、ＬＵ状態表４０８に記録された仮想ＬＵのアクセス先ＬＵを変更することにより、スナップショットを管理する。

【００７２】スナップショットの取得前、ＬＵ状態表４０８には、ＲＥＡＤ先ＬＵ、ＷＲＩＴＥ先ＬＵ共にミラー定義表のミラーＬＵが記録されている。スナップショット取得時、スナップショット管理プログラム４０６は、これらをミラー定義表に記録されているミラー元ＬＵに書き換える。また、スナップショット管理プログラム４０６は、バックアッププログラム４０５がアクセスするスナップショットの仮想ＬＵのＲＥＡＤ先、ＷＲＩＴＥ先ＬＵをミラー先ＬＵとし、状態を「スナップショット中」とする。

【００７３】例えば、図９及び図１０に示す状態で、仮想ＬＵ０のスナップショットを取得するとき、スナップショット管理プログラム４０６は、現在のアクセス先であるＬＵ０のミラー元ＬＵが、図９のミラー定義表を参照することによりＬＵ１であることが判るので、仮想ＬＵ０のＲＥＡＤ先ＬＵ、ＷＲＩＴＥ先ＬＵをＬＵ１に書き換え、状態を「スナップショット中」とする。また、スナップショット管理プログラム４０６は、仮想ＬＵ０のスナップショットとしてアクセスされる仮想ＬＵ１のＲＥＡＤ先ＬＵ、ＷＲＩＴＥ先ＬＵをＬＵ０のミラー先ＬＵであるＬＵ２に書き換え、状態を「スナップショット中」とする。これにより、仮想ＬＵ１へのアクセスが可能となり、また、仮想ＬＵ０、１の更新は、互いに二重化されないため、オリジナルデータ及びスナップショットのアクセスにそれぞれの仮想ＬＵを利用できるようになる。

【００７４】スナップショット管理プログラム４０６は、スナップショットを削除する場合、ＬＵ状態表４０８に記録されたスナップショットの仮想ＬＵの状態を「未定義」とし、オリジナルデータの仮想ＬＵの状態を「再同期中」に変更し、また、オリジナルデータの仮想ＬＵのＷＲＩＴＥ先ＬＵをミラー定義表に記録されているミラーＬＵとする。例えば、図９及び図１０に示す状態で、仮想ＬＵ２のスナップショットである仮想ＬＵ３

を削除するとき、スナップショット管理プログラム４０６は、仮想ＬＵ３の状態を「未定義」とし、また、仮想ＬＵ２の現在のＷＲＩＴＥ先ＬＵであるＬＵ４のミラーＬＵが図９のミラー定義表を参照することによりＬＵ３であることがわかるので、ＷＲＩＴＥ先ＬＵをＬＵ４からＬＵ３に書き換えると共に、状態を「再同期中」とする。これにより、スナップショットである仮想ＬＵ３へのプログラムからのアクセスが不可能となり、また、ミラー元ＬＵとミラー先ＬＵとの内容を一致させるデータコピー処理の際の、オリジナルデータの仮想ＬＵからの読み出しはミラー元ＬＵ、オリジナルデータの仮想ＬＵへの書き戻しはミラーＬＵがそれぞれアクセス先となる。

【００７５】データコピー処理が完了した後、スナップショット管理プログラム４０６は、オリジナルデータとしてアクセスされる仮想ＬＵのＲＥＡＤ先ＬＵを元のミラーＬＵに変更する。例えば、図９及び図１０の状態で、「再同期中」の状態にある仮想ＬＵ４のスナップショット削除を完了するために、スナップショット管理プログラム４０６は、図９のミラー定義表から現在のＲＥＡＤ先ＬＵであるＬＵ７のミラーＬＵがＬＵ６であることが判るので、ＲＥＡＤ先ＬＵをＬＵ６に書き換え、状態を「同期」とする。これにより、スナップショットの削除が完了する。

【００７６】前述した本発明の第２の実施形態によれば、第１の実施形態で説明した効果に加え、外部記憶装置をステートレス化したことにより障害時の回復処理の単純化を図ることができる。例えば、ＬＵの状態を変更する瞬間にコンピュータが障害を起こした場合、第１の実施形態の場合、回復処理においてコンピュータのＬＵ状態表に記録されている各ＬＵの状態とディスクアレイのミラー管理表に記録されている各ＬＵの状態を一致させる必要があるが、第２の実施形態の場合、ディスクアレイがステートレスであるため、コンピュータの状態のみを考慮した処理を行うことができる。

【００７７】図１２は本発明の第３の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図であり、以下、これについて説明する。この本発明の第３の実施形態は、データベースプログラムとバックアッププログラムとがそれぞれ別のコンピュータで動作する。この実施形態は、２つのコンピュータがＬＡＮで接続され、各コンピュータとディスクアレイとがファイバチャネルを用いた Storage Area Network(SAN) により接続され、また、バックアップ開始・終了を指示する管理コンソールとして別のコンピュータがＬＡＮで接続されて構成されている。

【００７８】図１２に示す第３の実施形態による計算機システムは、データベースサーバ１５１０、バックアップサーバ１５２０、管理コンソール１５４０が、それぞれＬＡＮインタフェース１５１３、１５２３、１５４１を介してＬＡＮ１５００により接続されて構成される。また、データベースサーバ１５１０とバックアップサーバ１５２０とは、ファイバチャネルインタフェース１５１４、１５２４を持ち、ファイバチャネルケーブル１５３１～１５３３とファイバチャネルスイッチ１５３０を介してディスクアレイ１５５０に接続されている。データベースサーバ及びバックアップサーバは、ファイバチャネルネットワークを経由してディスクアレイ１５５０内のデータにアクセスすることができる。ディスクアレイ１５５０の機能は、通信がファイバチャネル上に定義されるＳＣＳＩプロトコルに置きかえられた以外、第２の実施形態の場合と同一である。

【００７９】図１２において、データベースサーバ１５１０は、データベースプログラム１５１５、スナップショット管理プログラム１５１７、通信プログラム１５１６を実行し、バックアップサーバ１５２０は、バックアッププログラム１５２６及び通信プログラム１５２５を実行する。通信プログラム１５１６は、管理コンソールとＬＡＮで通信することによりバックアップ処理に関する指示受け取り、その内容に基づいてデータベースプログラム、スナップショット管理プログラムを呼び出す。同様に、通信プログラム１５２５は、管理コンソール１５４０からの指示に基づいてバックアッププログラム１５２６を呼び出す。データベースプログラム１５１５及びスナップショット管理プログラム１５１７の機能は、通信プログラムから指示を受けることができる点を除き第２の実施形態の場合と同様である。バックアッププログラム１５２６は、バックアップサーバ１５２０上にスナップショット管理プログラムがないため、仮想ＬＵではなくディスクアレイが提供する実際のＬＵを指定してアクセスを行う。

【００８０】管理コンソール１５４０は、バックアップ指示プログラム１５４２を実行し、ＬＡＮインタフェース１５４１とＬＡＮ１５００とを経由してデータベースサーバ１５１０及びバックアップサーバ１５２０上の通信プログラム１５１６、１５２５に指示を送る。

【００８１】次に、前述のように構成される本発明の第３の実施形態における管理コンソールとサーバとの通信について説明する。

【００８２】管理コンソール１５４０でユーザがバックアップの指示を出すと、バックアップ指示プログラム１５４２は、データベースサーバ１５１０の通信プログラム１５１６に対し、スナップショットを取得するように指示を行う。この指示を受信した通信プログラム１５１６は、データベースプログラム１５１５に対して、更新動作を一時停止するよう指示し、スナップショット管理プログラム１５１７にスナップショットの取得を指示する。スナップショット取得後、通信プログラム１５１６は、再びデータベースプログラムに指示を出してデータベースの更新を再開させ、スナップショット取得完了を

バックアップ指示プログラム１５４２に報告する。

【００８３】スナップショット取得完了の報告を受けたバックアップ指示プログラム１５４２は、次にバックアップサーバ１５２０の通信プログラム１５２５に対し、バックアップを開始するように指示を行う。この指示を受けた通信プログラム１５２５は、バックアッププログラム１５２６にバックアップ開始を指示し、バックアップを行わせる。バックアップ完了後、通信プログラム１５２６は、バックアップ完了をバックアップ指示プログラム１５４２に報告する。

【００８４】バックアップ完了の報告を受けたバックアップ指示プログラム１５４２は、通信プログラム１５１６に対してスナップショットの削除を指示する。この指示を受けた通信プログラム１５１６は、スナップショット管理プログラム１５１７に対して、スナップショットを削除するように指示を行う。スナップショット削除後、通信プログラム１５１６は、スナップショット削除完了をバックアップ指示プログラム１５４２に報告し、これにより、データベースのオンラインバックアップが完了する。

【００８５】前述した本発明の第３の実施形態によれば、第２の実施形態で説明した効果に加え、バックアップ中のデータベースサーバのＣＰＵ負荷の増大を軽減することができる。第２の実施形態は、バックアッププログラムとデータベースプログラムとが同一コンピュータ上で動作するため、バックアップ中のデータ転送処理にともなうＣＰＵ負荷増大により、データベースプログラムの実行速度が低下していたが、前述した本発明第３の実施形態は、バックアッププログラムがデータベースサーバとは別のコンピュータで動作するため、データ転送処理の負荷がデータベースサーバのＣＰＵ負荷に影響を与えることがなく、バックアップ中のデータベースプログラムの実行速度の低下を防止することができる。

【００８６】また、前述した本発明の第３の実施形態によれば、バックアップサーバとディスクアレイとの間がファイバチャネルで接続されているため、高速なデータ転送による迅速なバックアップを実現することができる。

【００８７】図１３は本発明の第４の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図、図１４はミラー定義表の構成を示す図であり、以下、これらの図を参照して第４の実施形態について説明する。この本発明の第４の実施形態は、更新前データの保存による従来技術のスナップショット管理方法から、前述した本発明の第１、第２の実施形態によるスナップショット管理方法に移行する場合に、従来のスナップショット管理プログラムの変更を最小限にとどめつつ本発明を適用し、スナップショット管理プログラムの大幅な変更を不要としたものである。

【００８８】ここで説明する本発明の第４の実施形態は、記憶内容が２つの異なる記憶媒体に二重化されるミラーＬＵにオリジナルデータを記憶し、スナップショット取得時点以後に更新が要求された場合に、更新前のデータを別領域に保存してからオリジナルデータを更新するものである。スナップショットは、スナップショット管理プログラムにより１つの仮想的なＬＵとして表されるが、更新されていない部分は、オリジナルデータを、更新された部分は、別領域に保存した更新前データを参照することによりアクセスされる。この第４の実施形態は、スナップショット管理プログラムの基本的な動作としては、更新前データの保存による従来の方法と同じであるが、ディスクアレイの支援機能と連携するために後述するアクセス先振り分け機能が追加されている。

【００８９】図１３に示す本発明の第４の実施形態におけるハードウェアの構成及びデータベースプログラム、バックアッププログラムの機能は前述で説明した本発明の第１、第２の実施形態と同様である。図１３において、ディスクアレイ４１０内で動作するミラー制御プログラム４１５は、前述した本発明の第２の実施形態の場合と同様に、ＬＵ間の関係を定義したミラー定義表１１１６を有している。ここでのミラー定義表１１１６は、その構造を図１４に示しているが、前述で説明した本発明の第２の実施形態における図９に示したミラー定義表とは異なり、ミラーＬＵとミラー先ＬＵとのみが定義される。第２の実施形態の場合と同様に、ミラーＬＵは、二重化された２つのＬＵを仮想的に１つのＬＵとして表したものであり、ミラー先ＬＵは、ミラーＬＵを構成する２つのＬＵの１つである。図１４に示す定義の場合、ＬＵ０はＬＵ１に二重化され、ミラー制御プログラム４１５は、ＬＵ０に対するアクセスコマンドを受け取ると、読み出しであればＬＵ０またはＬＵ１からデータを読み出し、書き込みであれば同じ内容をＬＵ０とＬＵ１との両方に書き込む。ミラー先ＬＵであるＬＵ１に対するアクセスは、ＬＵ１にのみアクセスし、ＬＵ０にアクセスすることはできない。

【００９０】また、ミラー定義表１１１６でミラー先ＬＵとして記録されているＬＵは読み出し専用であり、ミラー先ＬＵに対する書き込み要求は、ミラー制御プログラム４１５により拒否される。これにより、ミラー元ＬＵとミラー先ＬＵとの内容を常に一致させることができる。ミラーＬＵから読み出すデータとミラー先ＬＵから読み出すデータとの内容は同一であるが、ミラー先ＬＵにアクセスした場合、アクセス先のＬＵが１つに限定される点が異なる。データベースプログラム４０４等のコンピュータ上のアプリケーションプログラムは、常にミラーＬＵに対してアクセスを行う。ミラー先ＬＵは、後述するスナップショット管理プログラム４０６のアクセス先振り分け機能で使用するために提供される。

【００９１】スナップショット管理プログラム４０６は、オリジナルデータを参照するために従来の方法の場

合にミラーＬＵに対して発行されるアクセスコマンドの一部を、ミラー先ＬＵに振り分けるアクセス先振り分け機能を有している。オリジナルデータのミラーＬＵに対応するミラー先ＬＵは、ミラー定義表１１０８を参照することにより得られる。ミラー定義表１１０８は、ディスクアレイ４１０内のミラー定義表１１１６と同一の内容を持ち、第２の実施形態の場合と同様に、スナップショット管理プログラム４０６のMode Senseによるディスクアレイ４１０からの情報取得や、ユーザの手動設定によって作成される。

【００９２】次に、スナップショット管理プログラム４０６のアクセス先振り分け機能について説明する。

【００９３】オリジナルデータに対するアクセスは２種類に分けられる。その１つは、データベースプログラムからのオリジナルデータへのアクセスである。また、もう１つは、バックアッププログラムからのスナップショットへのアクセスであって、アクセス先のデータがスナップショット取得時点以後に更新されていない場合のアクセスであり、この場合、スナップショットが、オリジナルと記憶領域を共有しているのでアクセス先はオリジナルデータとなる。

【００９４】従来技術の方法をそのまま適用すると、前述のどちらのアクセスでも、オリジナルデータのアクセス用に提供されるミラーＬＵに対してアクセスコマンドを発行することになるが、本発明の第４の実施形態におけるスナップショット管理プログラム４０６は、後者のアクセス先をミラー先ＬＵに変更する。

【００９５】これにより、バックアッププログラム４０５から発生するオリジナルデータへの全ての読み出し要求は、そのアクセス先がミラー先ＬＵに限定され、従って、そのアクセス負荷がミラー先ＬＵが割り当てられたディスクに限定されることになる。

【００９６】一方、データベースプログラム４０４から発生する読み出し要求のアクセス先はミラーＬＵであり、二重化された２つのＬＵのどちらをも使用することができる。従って、この読み出し要求のアクセス先として、負荷の低いディスクを使用することにより、バックアッププログラムから発生するオリジナルデータに対する読み出しの負荷が高い場合でも、データベースプログラムのディスクアクセス速度が低下することはない。例えば、図１４に示すミラー定義表のＬＵ０、ＬＵ１が、それぞれ、図１３のディスク装置４２０、４２１に割り当てられている場合に、ＬＵ０のスナップショットを取得すると、バックアッププログラムからオリジナルデータへのアクセス要求は、全てＬＵ１をアクセス先とするためディスク４２１のみを使用する。データベースプログラム４０４から発生するアクセス要求のアクセス先はＬＵ０であるため、ディスク４２０、４２１のどちらでも使用することができる。このため、ミラー制御プログラム４１５が負荷の低いディスク４２０をアクセス先として選択すれば、バックアップによるディスクアクセス負荷が高い場合でも、データベースプログラムのディスク読み出し性能を低下させることがない。

【００９７】前述した本発明の第４の実施形態によれば、第２の実施形態の場合の効果に加え、従来技術のスナップショット管理プログラムを流用することができ、スナップショット管理プログラムの開発期間を短縮することができる。すなわち、第４の実施形態によれば、従来技術のスナップショット管理プログラムに、アクセス先振り分け機能を追加することにより、ディスクアレイの支援機能を利用するスナップショット管理プログラムを容易に作成することができる。

【００９８】前述した本発明の各実施形態によれば、オリジナルデータのアクセスに使用される記憶領域とスナップショットのアクセスに使用される記憶領域とが異なる記憶媒体上にあるため、バックアップ中の特定媒体への負荷の集中を避けることが可能となり、また、バックアップをオンライン処理とは別のコンピュータ上で実行することができ、これにより、バックアップ中のＣＰＵ負荷の増大を軽減することができる。

【００９９】また、本発明の各実施形態によれば、データ書き込みの二重化を外部記憶装置において行っているため、データ書き込みの二重化処理に伴うコンピュータのＣＰＵ負荷やデータ通信量の増加を生じさせることがなく、さらに、スナップショット取得中の更新位置の管理機能や、スナップショット削除の際に二重化された記憶内容の一致を保証する機能を持たない外部記憶装置を用いることができ、これにより、安価な計算機システムを構成することができる。

【０１００】

【発明の効果】以上説明したように本発明によれば、バックアップ中の特定媒体への負荷の集中を避けることが可能となり、また、データ書き込みの二重化処理に伴うコンピュータのＣＰＵ負荷やデータ通信量の増加を生じさせることを防止することができる。

【図面の簡単な説明】

【図１】本発明の第１の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図である。

【図２】ミラー管理表の構成を示す図である。

【図３】更新位置管理表の構成を示す図である。

【図４】Mode Select及びModeSenseで使用するパラメータページの内容を説明する図である。

【図５】スナップショット管理プログラムのデータ書き込み時の動作を説明するフローチャートである。

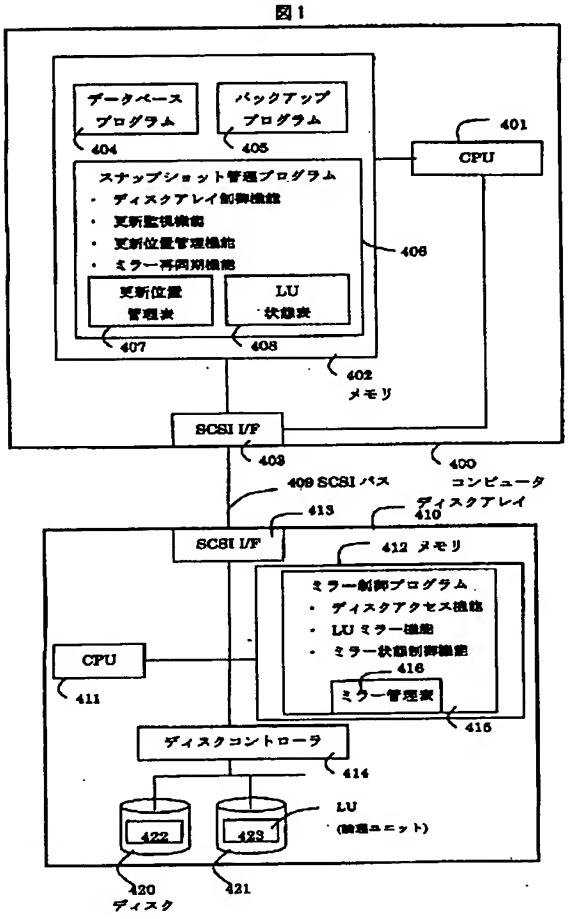
【図６】コンピュータからＷＲＩＴＥコマンドを受信したミラー制御プログラムの動作を説明するフローチャートである。

【図７】スナップショット管理プログラムのミラー再同期機能の動作を説明するフローチャートである。

【図８】本発明の第２の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図である。
【図９】ミラー定義表の構成を示す図である。
【図１０】ＬＵ状態表の構成を示す図である。
【図１１】ModeSense に対して返すパラメータページの内容を説明する図である。
【図１２】本発明の第３の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図である。
【図１３】本発明の第４の実施形態によるスナップショット管理を行う計算機システムの構成を示すブロック図である。
【図１４】ミラー定義表の構成を示す図である。
【符号の説明】
４００　コンピュータ
４０１、４１１　ＣＰＵ
４０１、４１２　メモリ
４０３、４１３　ＳＣＳＩインタフェース
４０４　データベースプログラム
４０５　バックアッププログラム
４０６　スナップショットプログラム
４０７　更新位置管理表
４０８　ＬＵ状態表
４０９　ＳＣＳＩバス
４１０　ディスクアレイ
４１４　ディスクアレイ
４１５　ミラー制御プログラム
４１６　ミラー管理表
４２０、４２１　ディスク装置
４２２、４２３　ＬＵ
１１０８、１１１６　ミラー定義表

【図１】

図1



【図２】

図2

415

LU	ミラー先 LU	状態
1	4	同期
2	5	スナップショット中
3	6	再同期中

【図１１】

図11

ページコード 00h
種別
ミラー元 LU
ミラー先 LU

【図３】

図3

407

LU	更新部開始 LBA	更新部終了 LBA
2	0	15
2	256	511
3	8	31

【図４】

図4

ページコード 00h
状態

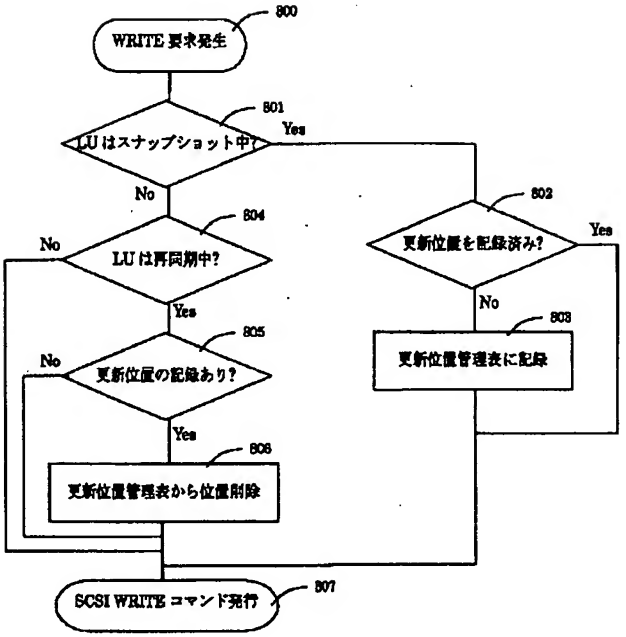
【図１４】

図14

ミラーLU	ミラー先 LU
0	3
1	4
2	5

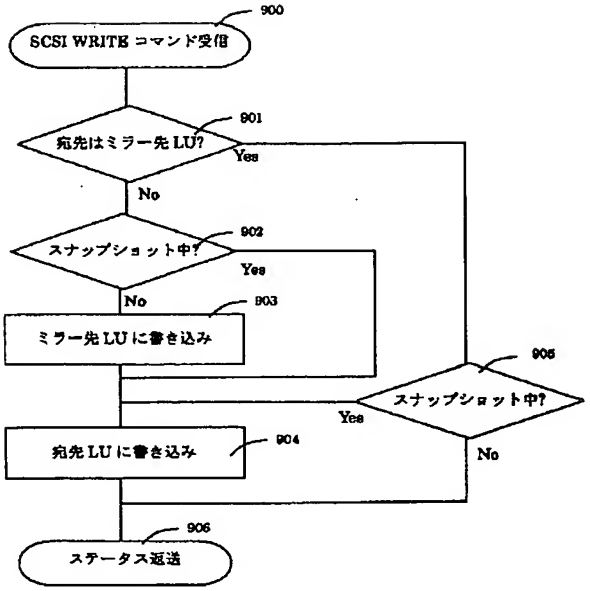
【図５】

図５



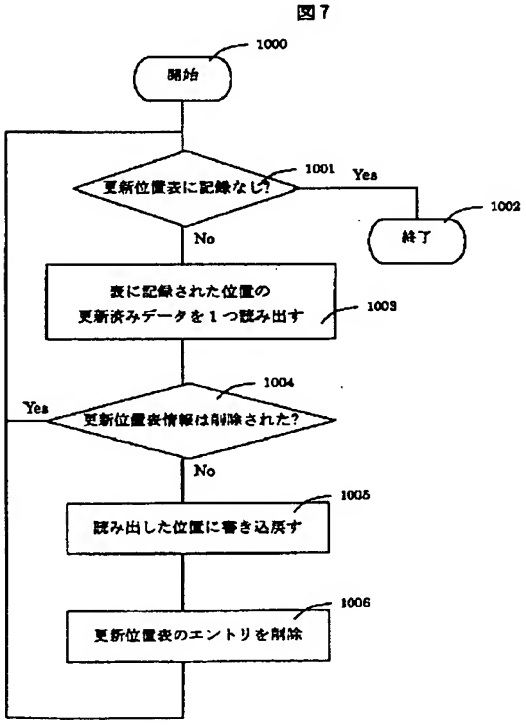
【図６】

図６



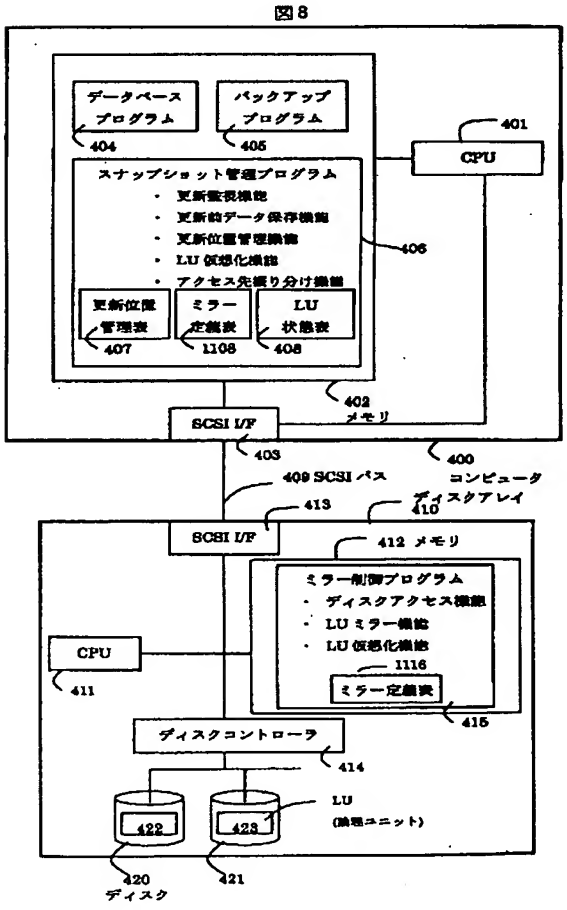
【図７】

図７



【図８】

図８



【図９】

図9

1116

ミラーLU	ミラー元LU	ミラー先LU
0	1	2
3	4	5
6	7	8

【図１０】

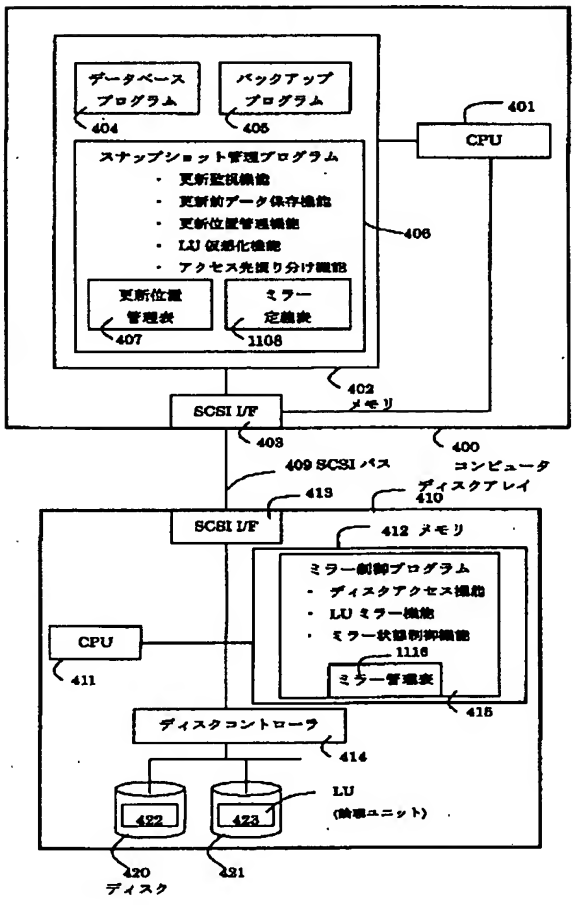
図10

408

仮想LU	READ先LU	WRITE先LU	状態
0	0	0	同期
1	—	—	未定義
2	4	4	スナップショット中
3	5	5	スナップショット中
4	7	6	再同期中
5	—	—	未定義

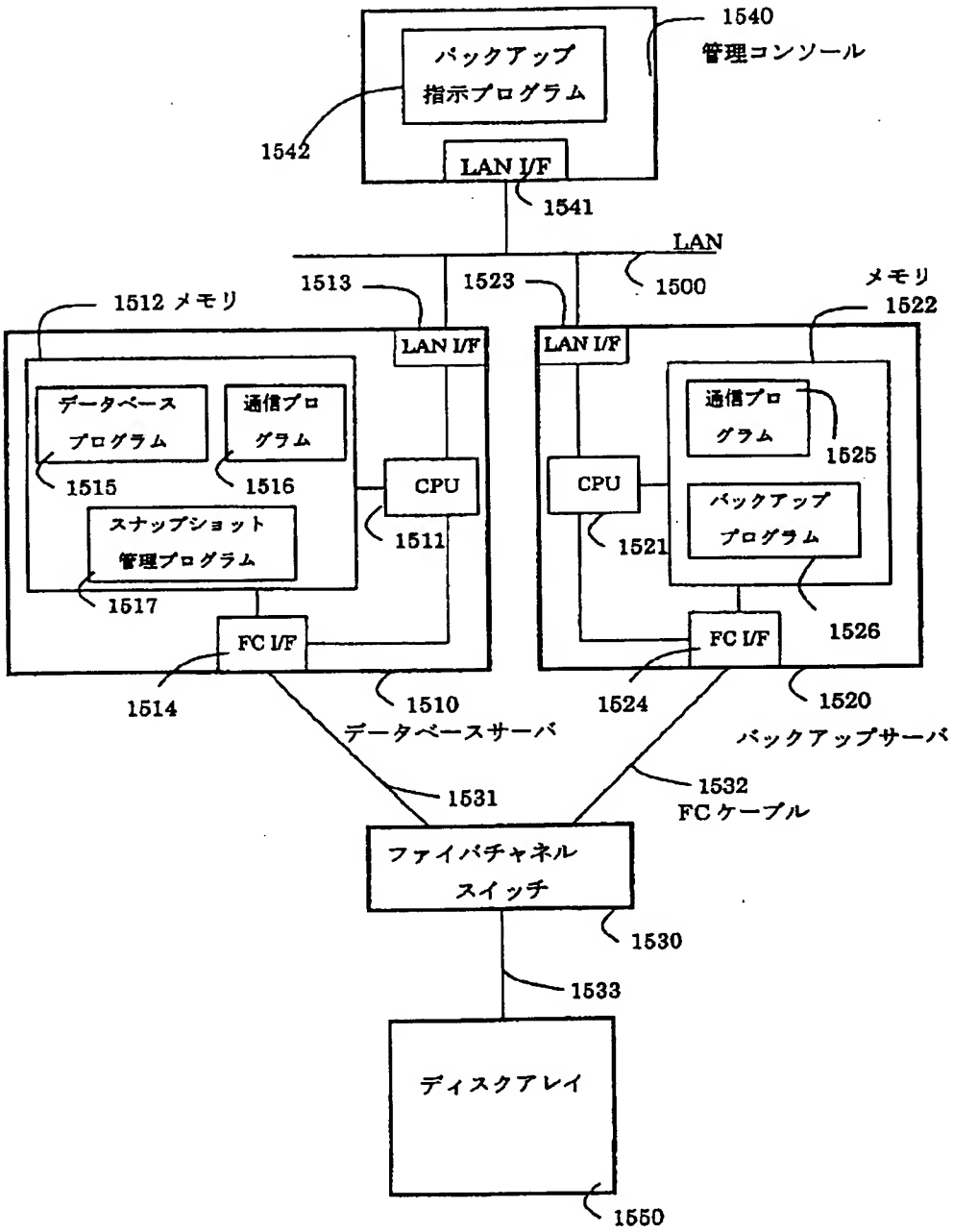
【図１３】

図13



【図１２】

図１２



フロントページの続き

(51) Int.Cl.[7]	識別記号	F I	ﾃｰﾏｺｰﾄﾞ（参考）
G11B 20/10		G11B 20/10	D

(72)発明者　村岡　健司
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内
(72)発明者　大枝　高
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(72)発明者　八木沢　育哉
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内

Ｆターム(参考)　5B018 GA04 HA03 MA12
5B065 BA01 CA30 EA02 EA12 EA34
5B083 AA08 AA09 BB01 BB03 CC04
CD11 EE08
5D044 AB03 BC01 CC04 DE03 DE12
DE72 DE92 HL02